SAPNet: Segmentation-Aware Progressive Network for Perceptual Contrastive Deraining
Supplementary Material

Shen Zheng, Changjie Lu, Yuxiong Wu and Gaurav Gupta
Wenzhou-Kean University
Wenzhou, China
zhengsh, lucha, yuxiongw, ggupta@kean.edu

1. Introduction

This supplementary material first displays the detailed model architecture for SAPNet, including the specific layers for the proposed progressive dilated unit (PDU) and channel residual attention (CRA). Next, we clarify the details about our user study and make a parameters comparison for different models. After that, we list more ablation study result in terms of qualitative and quantitative metrics.

Finally, we provide additional visual comparison on the performances of SAPNet against top-performing methods like DDN [1], RESCAN [5], PreNet [7], MSPFN [4], MPRNet [11], EffDerain [2], Syn2Real [10], and MOSS [3]. For synthetic rainy images, we select images from Rain100L [9] and Rain100H [9]. For real world images, we select images from MOSS [3] and SIRR [8]. Yolov3 [6] and PSPNet [12] is used for object detection and semantic segmentation, respectively.

2. Detailed Network Architecture

In this section, we supplement the detailed network Architecture for SAPNet. Specifically, we present the layer details about progressive dilated unit (PDU) in Table 1 and channel residual attention (CRA) in Table 2.

3. User Study and Model Parameter

3.1. User Study Comparison

We have conducted a user study for a comprehensive assessment of deraining methods on real-rainy images. Specifically, 100 adult participants from age 18 to 40 are randomly selected to attend the survey for the user study. We then randomly choose one photo from each of the three datasets, including Rain800, SIRR, and MOSS. The participants are asked to rate that three images with a score from 1 (worst) to 5 (best). The final user study score is the average from that three images. Some results of user study details are in Fig. 1. We can see that SAPNet outperforms others with the most '5' (47%) and the least '1' (2%)

3.2. Parameter Comparison

We draw the graph for model parameters and user study scores at Fig. 2. We find that SAPNet has a higher PSNR (29.46) and SSIM (0.897) and fewer model parameters (0.18M) than most methods. Although DDN, RESCAN and PreNet have a comparable model size, their PSNR and SSIM are significantly worse than SAPNet.

4. More Ablation Studies

4.1. Ablation of Limited Training Images

To demonstrate SAPNet's effectiveness at limited (labelled) training images, we conduct a quantitative comparison on Rain100H. Table 3 shows the performances of SAPNet when there is 40%, 60% and 100% training images. We can see that SAPNet maintains high PSNR and SSIM scores even without a large proportion of training images.

Require: rainy image y, recurrence step t, recurrent unit output x	
Name	**Details**
Add0	Add($\mathbf{y}$, $\mathbf{x}^{t-1}$);
Conv0	Conv(6,32,3,1,1); ReLU;
Res_Conv1	Conv(32,32,3,1,1,1); CRA(32,16); ReLU; Conv(32,32,3,1,1,1); CRA(32,16); ReLU;
Add1	Add(Conv0, Res_Conv1); ReLU;
Res_Conv2	Conv(32,32,3,1,1,2); CRA(32,16); ReLU; Conv(32,32,3,1,1,2); CRA(32,16); ReLU;
Add2	Add(Add1, Res_Conv2);ReLU;
Res_Conv3	Conv(32,32,3,1,1,4); CRA(32,16); ReLU; Conv(32,32,3,1,1,4); CRA(32,16); ReLU;
Add3	Add(Add2, Res_Conv3);ReLU;
Res_Conv4	Conv(32,32,3,1,1,8); CRA(32,16); ReLU; Conv(32,32,3,1,1,8); CRA(32,16); ReLU;
Add4	Add(Add3, Res_Conv4);ReLU;
Res_Conv5	Conv(32,32,3,1,1,16); CRA(32,16); ReLU; Conv(32,32,3,1,1,16); CRA(32,16); ReLU;
Add5	Add(Add4, Res_Conv5);ReLU;
Conv	Conv(32,3,3,1,1)
Output	$\mathbf{x}^t$

Table 1. The Architecture of Progressive Dilated Unit (PDU), where "Conv" is the convolution layer with (input channel, output channel, kernel size, padding size, stride, padding). CRA is channel residual attention block with (channel number, reduction factor). Here we omit convolutional LSTM for simplicity.

Require: Input x, Channel Number C, Reduction Factor R	
Name	**Details**
Pool	AdaptiveAvgPooling;
Conv_1×1	Conv(C,C,1,0,1);
Conv_du	Conv(C,C/R,1,0,1); ReLU; Conv(C/R,C,1,0,1);
Add1	Add(Conv_1×1, Conv_du); Sigmoid;
Output	Multiply(x, Add1)

Table 2. The Architecture of Channel Residual Attention Block (CRA), where "Conv" is the convolution layer with (input channel, output channel, kernel size, padding size, stride).

Metrics	40%	60%	100%
PSNR	26.49	27.37	**29.46**
SSIM	0.853	0.866	**0.897**

Table 3. Ablation result for SAPNet with limited training images

We also visually shows SAPNet' performance with limited training images in Fig. 3. To aid evaluation, we list the previous state-of-the-art MSPFN for comparison. It is shown that with 100% training images, SAPNet surpasses MSPFN in terms of rain removal ability. Even with 40% or 60% training images, SAPNet removes more vertical rain streaks and results in better contrast and exposure than MSPFN (100%). In short, SAPNet has great robustness at limited training images.

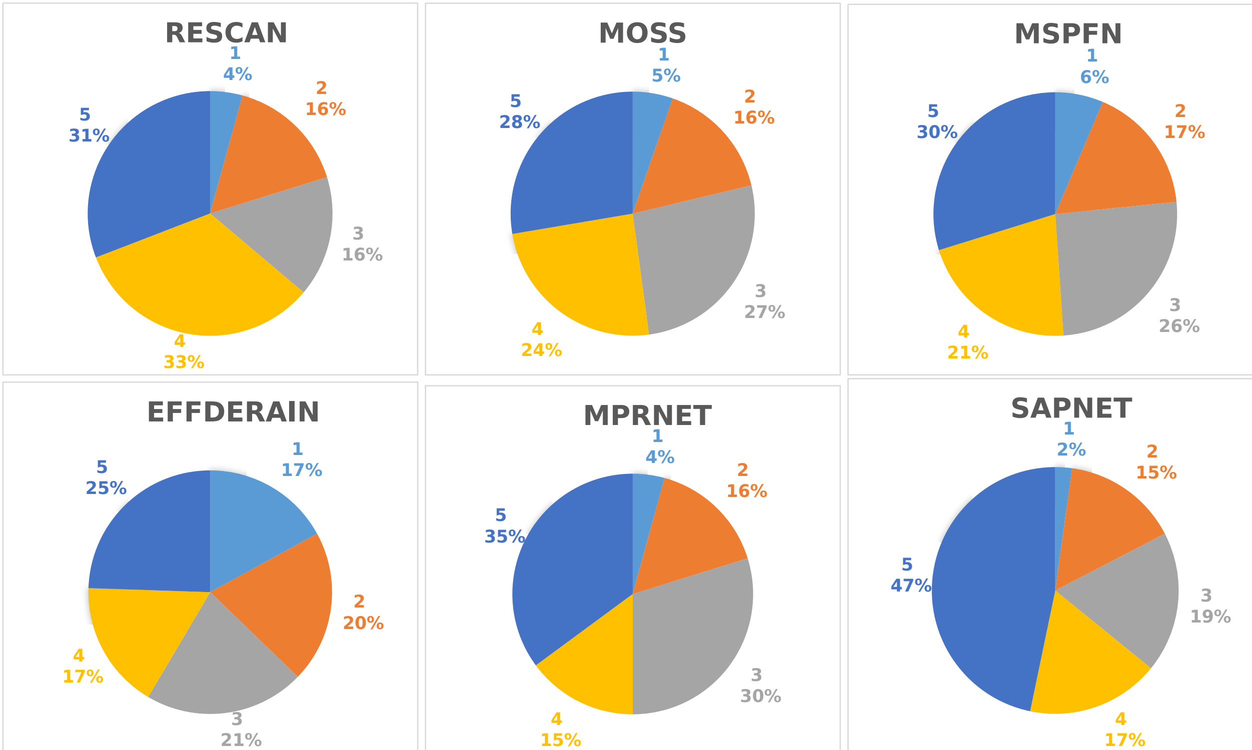
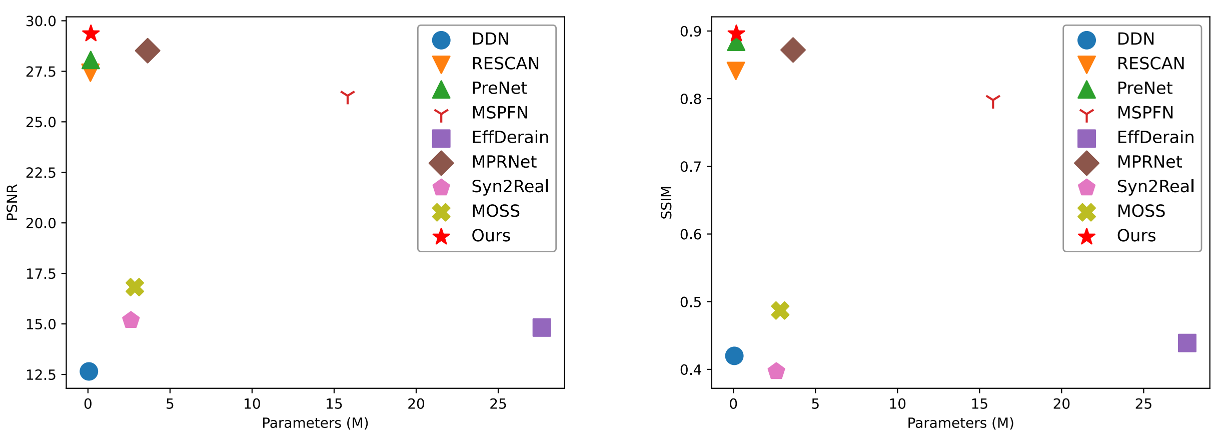


Figure 1. Detailed scores about the user study. A score of 1 indicates the worst performance, whereas a score of 5 indicates the best.



(a) PSNR-Param

(b) SSIM-Param

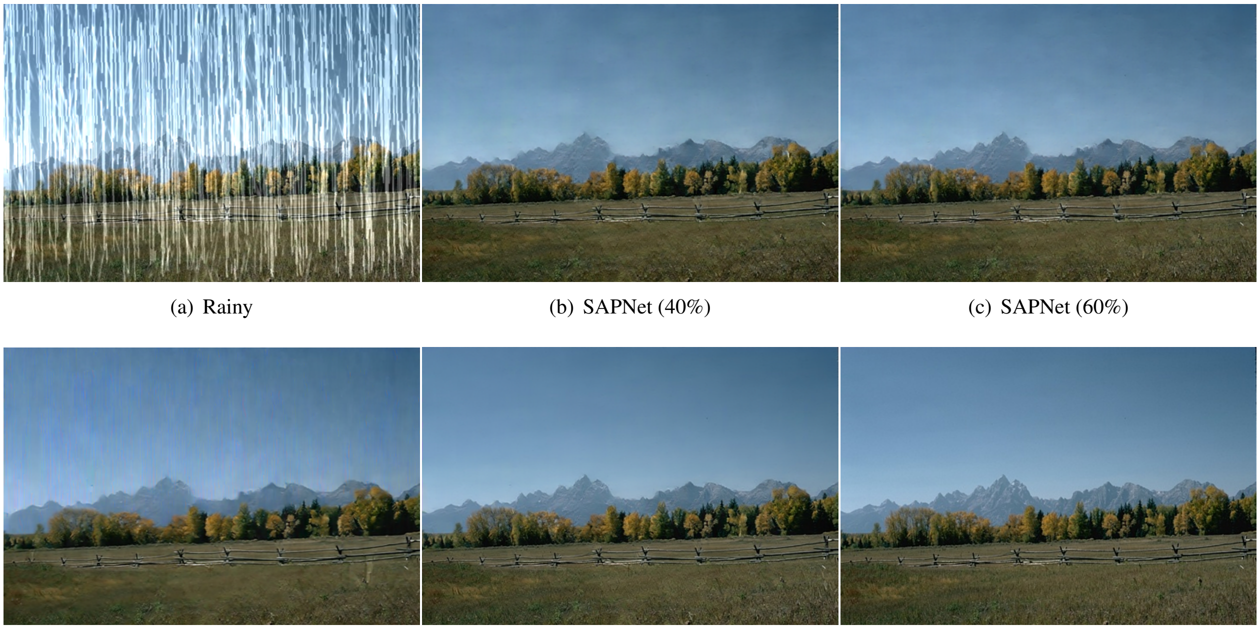
Figure 2. PSNR↑, SSIM↑ and Parameter↓ trade-off at Rain100H dataset

5. More Visual Comparisons

This section presents additional visual comparisons on the synthetic rainy dataset (Rain100H, Rain100L) and real rainy datasets (SIRR, MOSS). After that, we provide other results about semantic segmentation on the proposed CityScape150.

5.1. Synthetic Rainy Dataset

Fig. 4 displays the visual comparisons on the synthetic rainy dataset Rain100H. It is shown that SAPNet removes the most visible rain streaks and result in the most negligible amounts of artifacts.



(a) Rainy (b) SAPNet (40%) (c) SAPNet (60%)

(d) MSPFN (100%) (e) SAPNet (100%) (f) GT

Figure 3. Visual ablation of limited training data (Prop. %)

Fig. 5 shows the visual comparisons on the synthetic rainy dataset Rain100L. It appears that SAPNet is the only model that clear all large rain streaks and recover the details of the background mountains.

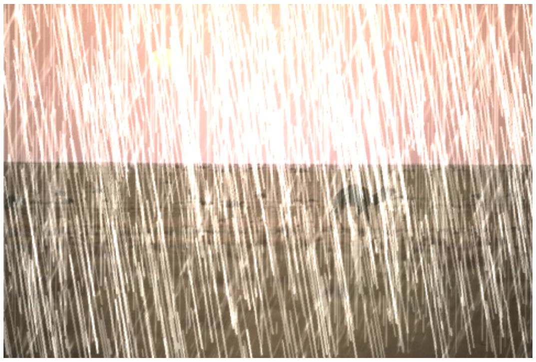
5.2. Real Rainy Dataset

Fig. 6 present the visual comparison on the real-world rainy dataset SIRR. We can see that SAPNet remove diverse types of rain streaks and preserve the bird's edge information.

Fig. 7 displays the visual comparison on the real-world rainy dataset MOSS. It is shown that SAPNet cleans the most amounts of rain streaks without distorting the illumination in the original rainy images.

5.3. Semantic Segmentation

Fig. 8 and Fig. 9 shows the semantic segmentation result comparison on the proposed CityScape150 dataset. Those two figures agree that SAPNet (1) preserves the most amount of semantic details (2) presents meaningful contents for segmentation (3) is the closest to the groundtruth.


(a) Rainy


(b) RESCAN


(c) PreNet


(d) MSPFN


(e) SAPNet (Ours)


(f) GT

Figure 4. Visual comparison at Rain100H dataset

(a) Rainy

(b) DDN

(c) MSPFN

(d) MPRNet

(e) SAPNet (Ours)

(f) GT

Figure 5. Visual comparison at Rain100L dataset

(a) Rainy (b) DDN

(c) RESCAN (d) PreNet

(e) MPRNet (f) SAPNet (Ours)

Figure 6. Visual comparison at SIRR dataset

(a) Rainy

(b) DDN

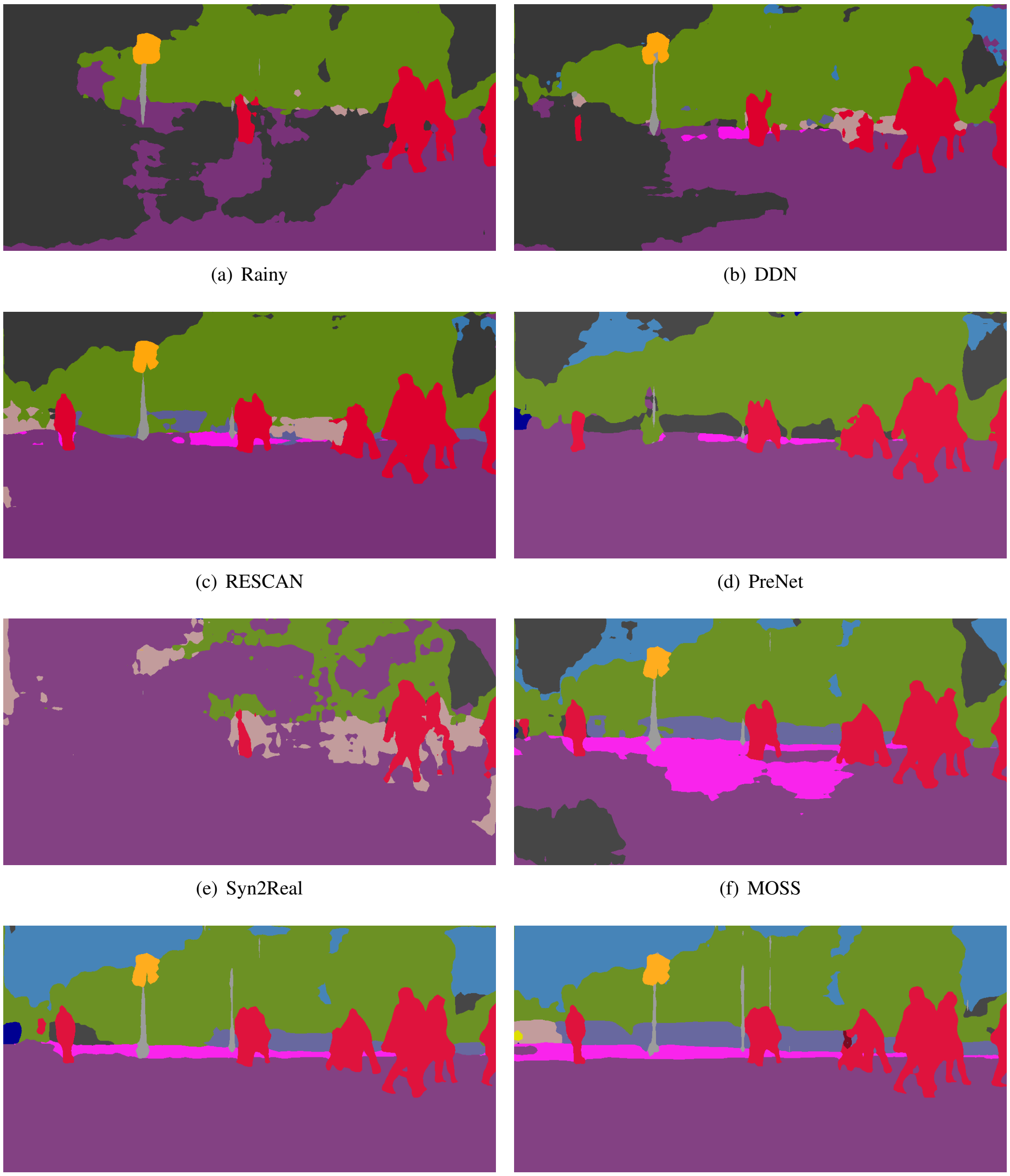
(c) EffDerain

(d) Syn2Real

(e) MPRNet

(f) SAPNet (Ours)

Figure 7. Visual comparison at MOSS dataset



(a) Rainy (b) DDN

(c) RESCAN (d) PreNet

(e) Syn2Real (f) MOSS

(g) SAPNet(Ours) (h) GT

Figure 8. Visual comparison at CityScape150 dataset (1)

(a) Rainy

(b) DDN

(c) RESCAN

(d) PreNet

(e) Syn2Real

(f) MOSS

(g) SAPNet(Ours)

(h) GT

Figure 9. Visual comparison at CityScape150 dataset (2)

References

[1] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3855–3863, 2017.

[2] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. *arXiv preprint arXiv:2009.09238*, 2020.

[3] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7732–7741, 2021.

[4] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355, 2020.

[5] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 254–269, 2018.

[6] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[7] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3937–3946, 2019.

[8] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, 2019.

[9] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017.

[10] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2726–2736, 2020.

[11] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[12] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.